DOCUMENT RESUME

ED 395 238 CG 026 983

AUTHOR Latkovich, Sharon A.
TITLE A Review of the Literature on the Validity of the Child Behavior Checklist.
PUB DATE Feb 96
NOTE 28p.; Paper presented at the Annual Meeting of the Eastern Educational Research Association (Cambridge, MA, February 21-24, 1996).
PUB TYPE Speeches/Conference Papers (150) -- Information Analyses (070)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Behavior Problems; *Diagnostic Tests; Elementary Secondary Education; Evaluation Methods; *Evaluation Research; *Interpersonal Competence; *Personality Assessment; Problem Children; *Psychological Evaluation; Psychological Patterns; Psychological Testing
IDENTIFIERS *Child Behavior Checklist

ABSTRACT

The Child Behavior Checklist (CBCL) is used to assess the behavioral problems and social competencies of children. Its broad usage in both a practitioner and research context has led to misapplications as compared to the CBCL's original intended purpose. Practical applications vary from the more traditional mental health centers and medical contexts to schools and forensic applications. Furthermore, an exhaustive number of citations of the CBCL can be found in the research literature. These 1990's uses of the CBCL most probably do not coincide with the original intent or purpose of the CBCL. Additionally, the composition of the normative sample presents particular difficulties for establishing validity across different groups. Thus given the broad usage of the CBCL combined with potential validity biases, it is the purpose of this study to review the literature regarding the validity of CBCL. This review examines the original purpose of the test, the normative sample, and test development from the purview of content, concurrent, predictive, and construct validity. A call is made for additional research so as to facilitate practitioners' and researchers' understanding of the nature of the CBCL in regards to a given use on a specific population. (Contains 43 references.) (Author/TS)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 395 238

Running Head: VALIDITY OF THE CHILD BEHAVIOR CHECKLIST

A Review of the Literature on the Validity of the Child Behavior Checklist

Sharon A. Latkovich

Ashland University

Presented at:

Eastern Educational Research Association

February 21-24, 1996

Cambridge, MA

BEST COPY AVAILABLE

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

S. LATKOVICH

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

CG026983

Abstract

The Child Behavior Checklist (CBCL) is used to assess the behavioral problems and social competencies of children. Its broad usage in both a practitioner and research context has led to misapplications as compared to the CBCL's original intended purpose. Practical applications vary from the more traditional mental health centers and medical contexts to schools and forensic applications. Furthermore, an exhaustive number of citations of the CBCL can be found in the research literature. These 1990's uses of the CBCL most probably do not coincide with the original intent or purpose of the CBCL. Additionally, the composition of the normative sample presents particular difficulties for establishing validity across different groups. Thus given the broad usage of the CBCL combined with potential validity biases, it is the purpose of this study to review the literature regarding the validity of the CBCL. This review will examine the original purpose of the test, the normative sample, and test development from the purview of content, concurrent, predictive, and construct validity.

A Review of the Literature on the Validity of the Child Behavior Checklist

The Child Behavior Checklist (CBCL) was developed by Achenbach and Edelbrock (1983) to "record in a standardized format the behavioral problems and competencies of children [ages 4 to 16] as reported by their parents or others who know the child well" (Kramer & Conoley, 1990, p. 36). The instrument is widely used by practitioners and researchers alike. It can be completed in a relatively short period of time (15 minutes) which most researchers and practitioners alike view as a distinct advantage (e.g., Mooney, 1984, Kelley, 1985; Christenson, 1990). Data is easily accessible given the use of parents as informants. The availability of a taxonomy of behavior profile types, based on a clinical sample, allows for stable comparative assessment (Mooney, 1984). Additional specific strengths include: "well written, informative, and very 'user friendly' manuals; comprehensiveness of the instrument in allowing professionals to gather standardized information from multiple sources; and the availability of norms for age groups of 6-11 years and 12-18 years by sex" (Christenson, 1990, p. 40-41).

Achenbach (personal communication, February, 1995) has claimed that "one million [CBCL test instruments] are used each year." Practical applications vary from the more traditional mental health centers and medical contexts to schools and forensic applications. Crawford (personal communication, July 16, 1995) used the CBCL fairly frequently when she was the sole school psychologist

in a county-wide system. Her rationale for such frequent usage was her estimation of the superior reliability of the instrument compared to the alternatives, and the CBCL's applicability to a wide age range (4-16). Kalverdijk (personal communication, August 1, 1995) also cited the broad age group as reasoning for his usage as a medical doctor in the Netherlands. Additionally, he mentions the computerized scoring program, the availability of "a large control datasample (sic)", and the implications of such wide usage for publication potential.

Furthermore, an exhaustive number of citations of the CBCL can be found in the research literature. For example, Kramer and Conoley (1990) included 115 test references to the CBCL in their review of the instrument for The Supplement to the Tenth Mental Measurements Yearbook. Perrin, Stein, and Drotar (1994) have referred to the CBCL as "the gold standard" in behavioral research on children; they refer to the tremendous number of citations of the CBCL in the Journal of Pediatric Psychology. This extensive research is partially attributable to the availability of a taxonomy of behavior profile types, which allows for stable comparative assessment (Mooney, 1984). This set of normative behavior profiles was derived from a clinical sample, the racial breakdown of which was 81.2% white, 17.1% black, and 1.8% other races. Informants were 83% mothers, 11.5% fathers, and 5.6% other respondents (Mooney, 1984, p. 177).

Given this broad usage of the CBCL in both a practitioner and research context, it is the purpose of this study to review the literature as to the appropriate use of the CBCL. In other words, this study will examine the validity of the CBCL. The CBCL is so frequently used that it has been assumed to be appropriate (e.g., Crawford, personal communication, July 16, 1995; Kalverdijk, personal communication, August 1, 1995). This review of the literature will examine the CBCL's measurement of psychosocial constructs and identify its appropriate and inappropriate uses.

Validity

Validity is the extent to which a test measures what it is intended to measure (Jensen, 1980; Wiersma & Jurs, 1990; Newman & Newman, 1994). It is that intent, the use of a test or the inferences derived from its results, that is the essential component of whether an instrument is well-grounded and appropriate. Barnett and Zucker (1990) concur that the key to validity is the test's usefulness for a specific purpose, or what Barrios and Hartmann (as cited in Barnett & Macmann, 1990) refer to as problem identification. Thus, validity must be examined in light of the original intended purpose of the instrument, the normative population on which the test was developed, and the actual procedures used by test developers (Barnett & Zucker, 1990, p. 59).

Original Purpose

For mental measurement testing in general, the first evaluative criteria, the original intended purpose, was the amelioration of deplorable social and living conditions at the turn of the twentieth century. Even then, there was a struggle for a balance between the best and most valid measures and the associated cost issues. While adaptability to a harsh environment remained a cornerstone of behavioral assessment, the purpose of testing was continually influenced by the corresponding evolution in childhood developmental theory.

In the 1960s, the hotbed of such theoretical debate revolved around the view of adolescent development as characterized by volatility and turmoil (e.g., Freud, 1965; Reed & Sautter, 1990). This view of adolescent development "prompted important new research from the scholarly community in the mid-1960s" (Powers, Hauser, & Kilner, 1989, p. 200). At approximately the same time, the American Psychological Association (APA) had as yet failed to rectify the glaring omission of diagnostic criteria for children. The concurrence of research on the volatile adolescent along with improving the reliability and validity of children's' diagnoses gave rise to "a 1966 study by Achenbach in which he 'content analyzed' over 600 clinical case histories of children " (Mooney, 1984, p. 168). This study served as the basis for the final 118 problem items in the CBCL. Thus, the original intended purpose or problem identification of the CBCL arose out of the volatility and turmoil school of adolescent development along with the lack of diagnostic aids for children.

However, Powers et al. (1989) refute the dark view of adolescence; "An important difficulty with the 'storm and stress' view of adolescence is that it blurs the boundaries between normality and pathology during adolescence" (p. 201). The implication is that the measurement of the CBCL's content circa the 1960's is not uniformly held today. As Barnett and Zucker (1990) point out, validity is a dynamic concept because of the centrality of usage; in other words, validity needs to be retested for each new use or population. Too often, practitioners (and some researchers), such as Kalverdijk (personal communication, August 1, 1995), use the CBCL based on the assumption of a well-validated instrument without a clear comprehension of the validity for the specific usage involved. Drotar, Stein, and Perrin (1995) note that this scenario of increased usage is a potent prelude to "misapplication and/or misinterpretation of data" (p. 185).

Normative Sample

The second validity criteria is the normative population on which the test was developed. This is a particular issue for the CBCL. The behavior problem scales of the CBCL was derived from a clinical sample of 2,300 children from 42 eastern mental health and related service agencies (Mooney, 1984, p. 169). Then, principal components analysis was used to derive the narrow-band scales, by sex and age. A second-order factor analysis was also run to determine the broad-band scales of internalizing and externalizing.

As a result of using a referred sample for the development of norms, the behavior of non-referred children for whom the CBCL was designed was not evaluated, a point noted by Perrin, Stein, & Drotar (1991). They specifically noted that the CBCL's focus on the identification of abnormal behavior was not conducive to the increased need for research on variations in normal childhood behavior, a contention also made by Emerson, Crowley, and Merrell (1994). Indeed, Achenbach (1991b) himself found minimal variance explanation in the Activities and Somatic Complaints scales; he suggests using these scales more in a descriptive vein owing to their insignificant concurrent validity (p. 94).

By comparison, the normative population for the social competence scales was comprised exclusively of non-referred children, at least 72% of whom were white children. In 80% of this sample, the mother was the respondent (Achenbach, 1991b, p. 22). The development of the social competence items was based on extensive pilot testing. So, the CBCL norms for social competence reflect "normal" capabilities evaluated on multiple draft test items while the norms for behavior reflect "abnormal" functioning using factor analysis.

Also, proper test administration and accurate scoring are necessary conditions for validity, as opposed to the limited diversity in the normative population which lowers validity. "The validity coefficients tend to decrease as the groups become more homogeneous" (Wiersma & Jurs, 1990, p. 201). Thus the more similarity in a sample, such as one racial group or a given socio-economic status (SES) cluster, the more limited the generalizability of the results.

Test Development

The third validity criteria centered on the development of the instrument itself. Good, thorough test development will enhance validity; the alternative is ambiguous questions or other test characteristics that detract from validity. In general, increasing test length with additional items of similar content will also increase validity.

Herein lies the strength of the CBCL as Achenbach and his colleagues relied on the statistical sophistication of second-order factor analysis thereby reducing the errors inherent in the subjective adjudgment of behavioral assessment or content. By the use of second-order factor analysis, higher-order patterns of functioning can be identified that would be obscured by a more simplified factor analysis; consequently, potential content measurement errors are avoided. For example, the research of McConaughy, Achenbach, and Gent (1988) confirmed better cognitive, academic, and social functioning among internalizers than externalizers, in addition to important differences between profile types within these two groups (p. 506).

Unfortunately, this statistical strength is somewhat offset by the incomplete evaluation of the social competence scales. Emerson et al. (1994), Mooney (1984) and Perrin et al. (1991) have all commented on the less than comprehensive nature of the competence scales, with Perrin et al. specifically pointing to the CBCL's measurement of accomplishments and participation as opposed to capacity or facility.

Content Validity

Of Barnett and Zucker's (1990) three criteria, content validity aligns most closely with the first criteria, intended purpose. More specifically, content validity measures how representative the test items are of the objective area which the test is attempting to measure (Newman & Newman, 1994; Barnett & Zucker, 1990; Wiersma & Jurs, 1990; Jensen, 1980). Therefore, the content validity of the CBCL involves assessment of the prospective purpose in light of Achenbach's original purpose.

Evaluation of content validity also includes expert judgment, a method commonly known as face validity. This methodology involves a subjective analysis of the correspondence between the test items and the content being assessed; it is dependent on the specific definition of the content and its related elements. The test items must be adjudged to represent a sufficiently broad coverage of the content area. High face validity indicates a distinct portrayal of the behavior being assessed, and therefore, answers are quite easy to "fake" through biased or distorted responses (Bornstein et al., 1994). Additionally, face validity does not generate an objective measure, such as a correlation coefficient.

Herein lies the criticism of content validity. "Face validity is the Rodney Dangerfield of psychometric variables: It has received little attention -- and even less respect – from researchers" (Bornstein, Rossner, Hill, & Stepanian, 1994, p. 363). Shepard (as cited in Reynolds, 1995) points out that there is minimal support that "anyone can – upon surface inspection -- detect the degree to which

any given item will function differentially across groups" (p. 556). Indeed, Elliot, Busse, and Gresham (1993) indicate that content validity is not necessary for empirically-derived scales (such as the CBCL) as these syndromes are defined statistically. Yet those who do endorse the use of face validity champion its public relations value. The perception of validity in the eye of the public is not something to be taken lightly (Reynolds, 1995, p. 556, 557).

Likewise, the perception of content bias is also an important consideration. Content bias has been defined as: "An item or subscale of a test is considered to be biased in content when it is demonstrated to be relatively more difficult for members of one group than for members of another" (Reynolds, 1995, p. 553). Therefore, one test of content validity is a group-by-item interaction; if the test is valid, this interaction term should be insignificant. Reynolds notes, however, that occurrences in the literature of race-by-item interaction have accounted for a small proportion of variance. As a result, elimination of those items will have a negligible effect on ameliorating content validity.

An example of Reynolds' operational definition of content bias in the CBCL is found in the work of Perrin, Stein, and Drotar (1991) and Raadal, Milgrom, Cauce, and Mancl (1994). Perrin et al. found that children with chronic physical disorders would have inappropriately elevated scores, especially on the Somatic Complaints subscale where the majority of these symptoms are scored.

Raadal et al. noted that the validity of the CBCL norms for low socioeconomic status (SES) children was questionable.

Concurrent Validity

A second type of validity is concurrent validity. Concurrent validity estimates validity based on the correlation between the test instrument and another test that has already had its validity estimated (Newman & Newman, 1994, p. 53). Hence, it is inferred that since the current test, the CBCL, and the previously validated test are correlated, the current test is also correlated with the criterion which as Jensen (1980) indicates is a risky assumption. Witt, Heffer, and Pfeiffer (1990) cite the significant correlation of the CBCL with the Conners Parent Rating Scale and the Revised Problem Checklist (p. 369). Achenbach, Howell, Quay, and Conners (1991) found concurrent validity between the CBCL and a new instrument, the ACQ Behavior Checklist which was named for its key developers, Achenbach, Conners, and Quay.

To the preceding definition of concurrent validity, Jensen (1980) specifies that concurrent validity also refers to "the correlation between a test and a criterion when both measurements are obtained at nearly the same point in time" (p. 301). Various criterion measures can be used, such as behavioral classifications or Diagnostic and Statistical Manual (DSM) categories of disorders. However given the historical minimal classification of children by the DSM, the concurrent validity of an instrument such as the CBCL as compared to DSM categories may not be meaningful. Thus instead of using the DSM as a

criterion, Achenbach's (1991c) research uses referral/non-referral for mental health services as a criterion for testing the concurrent validity of the individual CBCL scales, a study which Kelley (1985) refers to in her approbatory review of the validity of the CBCL. Yet despite his obvious negative opinion of the DSM, Achenbach (1991b) cites "studies which show significant relations between DSM diagnoses and pre-1991 CBCL scores" (p. 88).

Furthermore if the criterion is a diagnosis, then validity is also dependent on the accuracy of the diagnostic assessment, a point made by Chen, Faraone, Biederman, and Tsuang (1994). Their research looked at the concurrent validity of the CBCL with a diagnosis of Attention-Deficit Hyperactivity Disorder (ADHD), finding that the Attention Problems scale was the best predictor of ADHD in their samples. These results mirror those of Achenbach (1991b) who found the Attention Problems scale to have strong concurrent and predictive validity with referral status.

However, Chen et al.'s samples were composed of white male children, while Achenbach's (1991c) non-referred sample was 74% white, compared to 83% white for the referred sample, and the sample employed by Emerson, Crowley, and Merrell (1994) was 95% white. Both Chen et al. and Emerson et al. noted the lack of generalizability to non-white children and issued a call for additional research on different ethnic groups; Achenbach, on the other hand, found few effects of race; however, this may be due potentially to the homogeneous racial composition of the sample. Additionally, graphical

representations of his findings indicate disordinal interaction between referral status and the child's sex for Activities, Social, Total Competence, Anxious/Depressed, Social Problems, and Thought Problems, yet interaction with the variable sex was not specifically tested.

Predictive Validity

Predictive validity is closely related to the criterion conceptualization of concurrent validity, but the difference is timing. Whereas concurrent validity is based on the relationship between the test and a criterion measured at or about the same time, predictive validity applies if there is an intervening period (Wiersma & Jurs, 1990, p. 189). More specifically, predictive validity is the ability to predict a future outcome(s) significantly better by the use of the test than by mere chance alone (Newman & Newman, 1994, p. 53). As Barnett and Zucker (1990) note, the real test of predictive validity is "an analysis of actual outcomes. If better decisions are made as a result of including the measure, the test possesses predictive validity" (p. 66). This comment is suggestive of Sechrest's (as cited in Barnett & Zucker, 1990) "incremental validity:" the importance of the differential contribution of the test instrument (p. 67). It is this quality that led Most and Zeidner (1995), Barnett and Macmann (1990), and Jensen (1980) to label predictive validity as one of the most important types of validity in regards to the practical use of psychological tests.

According to Elliot, Busse, and Gresham (1993), the CBCL does possess this elusive quality of predictive or incremental validity, "despite some

shortcomings" (p. 317). For example, Verhulst, Koot, and Van der Ende (1994) conducted a longitudinal study of the CBCL and found that "total problem scores in the deviant range on the CBCL were significantly associated with poor outcomes six years later." (p. 531). Verhulst was associated with the Dutch translation of the CBCL. Their results confirmed the concurrent validity of the Attention Problems scale.

The practical importance of predictive validity infers a similar practical significance towards potential bias. Bias in predictive validity refers to the lack of random error in prediction (Reynolds, 1995). A factor that can yield misleading predictive validity is the definition of the criterion. For example, certain forms of behavior may be invulnerable to a specific definition; thus, a poor or vaguely defined criterion can act as an impediment to predictive validity. Also, variables other than the test or the criterion can affect the predictive ability of the test, such as the environment, race, gender, or socioeconomic status. For instance if the predictive measure shows a distinct pattern by different racial groups, it could be concluded that the measure was biased.

Moran (1990) reviewed this type of research that investigated cultural differences in performance on objective tests. Her findings "indicate that minority groups, on the average, earn more deviant scale elevations than do Anglos. . . . Blacks are generally found to have elevated schizophrenia and

hypochondriasis" (p. 531). The mere fact that differences exist does not equate to bias, but bias does lead one to question validity.

Accordingly, Moran called for further study as the quality of existing studies was found to be lacking and research on the predictive validity of instruments used with children by race almost non-existent. The wide use of the CBCL combined with this paucity of research begets the need for a study of the predictive validity of the CBCL for different racial groups.

Construct Validity

"Construct validity is a conglomeration of all other types of validity" (Newman & Newman, 1994, p. 54). "It encompasses both the criterion and content analyses, as well as providing a more theory-based evaluation of the logical and empirical bases of determining how well a score represents a construct. Several authors have suggested a more unified view of validity in which the role of construct validity is fundamental" (Tittle, 1994, p. 6316).

Thus, construct validity concentrates on an evaluation of the original theoretical underpinnings or constructs of the test instrument. Additionally, the instrument must measure the theoretical construct the same for each group on which it is used (Moran, 1990). Further compounding the issue is the fact that those constructs are hypothetical in psychological testing; in other words for any given construct, there can be uniform agreement or legitimate disputes. As a result, an evaluation of construct validity must be preceded by a thorough analysis of the particular theoretical point of view of the test developer(s), i.e.,

the original intended purpose. However, this purpose(s) "may be tangential to professional practice issues" (Barnett & Zucker, 1990, p. 64).

A subset of this definition of construct validity is convergent and divergent (or discriminant) validity. Convergent validity is the positive correlation between a test and criterion. Achenbach (1991b) claims that there is evidence of convergent validity between the CBCL and the DSM approach, despite his obvious loathing for the APA-designed categories. Epkins and Meyers' (1994) own empirical research indicated that the convergent validity of the CBCL differed by the child's gender in addition to finding an informant effect in the convergent validity research. Jensen, Traylor, Xenakis, and Davis (1988a) hypothesized that the informant effect might reflect low convergent validities between the scales. Emerson, Crowley, and Merrell (1994) also tested the construct validity of the CBCL with the School Social Behavior Scales (SSBS), finding convergent validity for the social competence scales ($r = .31$ to $.39$). However, Drotar, Stein, and Perrin (1995) disagree with these findings, noting the skewness of the social competence scores in non-referred children.

Divergent or discriminant validity, on the other hand, is a negative correlation between a test and a criterion. Emerson et al. (1994) found discriminant validity for the problem behavior scales ($r = -.30$ to $-.37$). More specifically, McConaughy, Mattison, and Peterson (1994) tested the discriminative validity of the CBCL for differentiating children with serious emotional disturbance (SED) from children with learning disabilities (LD),

finding that SED children scored significantly higher than those with LD on all scales except Somatic Complaints; the most significant predictors were the Thought Problems and Delinquent scales. Fonhomme (1992) found that the CBCL discriminated between clinical and non-clinical French children when age, SES, and sex were controlled for. Macmann, Barnett, and Lopez (1993) hypothesize that "the high degree of correlational overlap across the Attention Problems and Aggressive Behavior scales can be traced in part to discriminant validity problems at the item level" (p. 323). "Unfortunately, the discriminant validity research in childhood syndromes is limited" (Epkins & Meyers, 1994, p. 365).

In contrast to this observation of limited discriminant (construct) validity research, Kelley (1985) refers broadly to the research support of the CBCL's construct validity, yet she does not mention any specific citations. However, Mooney (1984) rectifies this error, citing the work of Weissman, Orvaschel, and Padian; Hodges, McKnew, Cytryn, Stern, and Klein; Hazzard, Christensen, and Margolin; Last and Bruhn. Of these studies, most dealt with the construct validity of the total behavior problem score of the CBCL. Barnett and Zucker (1990) also mention the importance of a clear link between the factor-derived syndromes and specific intervention strategies (p. 64).

Factor Analysis

A statistical technique that is commonly used to assess construct validity is factor analysis. By analyzing the intercorrelations between various items or

variables, factor analysis "reveals the major dimensions that underlie a set of items" (Barnett & Zucker, 1990, p. 61). "This is a form of construct validity, because factors may be viewed as theoretical constructs used to explain the sources of individual differences in a variety of psychological measurements" (Jensen, 1980, p. 304). Especially in psychological testing where the volume of data and the potential number of factors is overwhelming, it has been suggested that principal components analysis is used first to describe the underlying theoretical constructs (Stevens, as cited in Barnett & Zucker, 1990).

Following this logic, Achenbach and Edelbrock used principal components analysis to develop the narrow band syndromes of the CBCL. The first order factors, which comprised the narrow band syndromes, were specified as having at least five items with minimum loadings (correlation between a factor and the item score) of .30. The patterns of the loadings on an individual factor are the basis for the factor's description.

However, the normative sample was based on a population of clinically referred children which could potentially jeopardize the CBCL's validity in certain situations (Jensen et al., 1988a). Also, the replication studies "are most often based on boys: important sex differences are evident for certain syndromes" (Barnett & Zucker, 1990, p. 63). Consequently, the stability and replicablity of the factor structure might be questionable. Therefore, Emerson et al. (1994) issued a call for further research on this issue of factor structure stability.

Factor analysis can also be used to examine construct validity across groups by a comparison of factor structures. If the constructs are measured in the same way and with equal accuracy, the test will have been shown to exhibit construct validity. Reynolds (1995) notes that "consistent factor analytic results across populations do provide strong evidence that whatever is being measured by the instrument is being measured in the same manner and is, in fact, the same construct within each group" (p. 559).

However if the factor structures differ, then construct validity does not exist and it would be inappropriate to give the same theoretical interpretation to both groups. In fact, this is the definition of bias in construct validity: when a test measures different constructs for different groups (Reynolds, 1995). For example, the same construct can lead to different linguistic operationalizations, depending on the culture; certain psychological constructs are culture-specific (Moreland, 1990; Drotar et al., 1995). As a result, a compelling case can be made for the study for various race and ethnic groups so as to analyze differential construct validity (Moran, 1990; Emerson et al., 1994; Drotar et al., 1995). Drotar et al. (1995) even go so far as to comment that while the normative sample was chosen to represent the overall U.S. population, the norms are not equally applicable when applied to those children who are underrepresented in the sample, a point also made by Hill, Billingsley, Engram, Malson, Rubin, Stack, Stewart, and Teele (1993) in reference to the black community. Reynolds (1995), however, found:

> A large number of popular psychometric assessment instruments have been investigated across races and genders with a variety of populations of minority and white children. . . All roads have led to Rome: No consistent evidence of bias in construct validity has been found with any of the many tests investigated" (p. 563).

Yet, the "many tests" nor the research citations are not provided.

A key to validity is the link between the original intended purpose and its specific, research or practical, use.

Achenbach and McConaughy (1987) would concur that validity is a central concern, yet they suggest that standardized assessment and covarying sex and age are sufficient criteria to avoid biases in validity. Samuda (as cited in Algozzine, Wong, & Obiakor, 1994) would have responded that "standardized tests 'preserve the status quo' and relegate 'Blacks and other minorities to an inferior position in the larger society' " (p. 716). Algozzine et al., thus, issued a call for research into the validity of standardized tests for minorities.

Conclusion

The inherent difficulties of validity in psychological instruments is reflected in the myriad of definitions and operating criteria used to make diagnoses. This discussion has now come full circle back to the definition of validity: the intent to which a te.. measures what it is intended to measure (Jensen, 1980; Wiersma & Jurs, 1990; Newman & Newman, 1994). Too often, practitioners use the CBCL based on the assumption of a well-validated

instrument without a clear comprehension of the validity for the specific usage involved.

This assumption of equating wide usage with validity is exacerbated by the homogeneity of the normative samples which in turn limits the generalizability of any results. The norms are not equally applicable when applied to those children who are underrepresented in the sample. Achenbach and McConaughy's (1987) response that standardized assessment and covarying sex and age are sufficient to avoid validity biases conflicts with Samuda's (as cited in Algozzine et al., 1994) contention that standardized tests "preserve the status quo." Covariates found to impact the validity of the CBCL include age, sex, informant, SES, and race

The strength of the CBCL lies in the second-order factor analysis used on the behavioral problems scales. Yet, even this strength is questioned given the demographic characteristics of the normative sample. Again, research which would compare factor structures across groups would provide the support of the CBCL's construct validity, or lack thereof. Likewise, limited investigation of the aforementioned covariates and any potential interaction also begets the need for research.

This call for research should not be interpreted as a condemnation of the validity of the CBCL. Instead, it should signal recognition for practitioners and researchers alike to understand the nature of the CBCL in regards to a given use on a specific population.

References

Achenbach, T.M. (1985). Assessment and taxonomy of child and adolescent psychopathology. Beverly Hills, CA: Sage Publications.

Achenbach, T. M. (1991a). Integrative guide for the 1991 CBCL/4-18, YSR, and TRF profiles. Burlington, VT: University of Vermont Department of Psychiatry.

Achenbach, T. M. (1991b). Manual for the Child Behavior Checklist/4-18 and 1991 profile. Burlington, VT: University of Vermont Department of Psychiatry.

Achenbach, T. M., Howell, C. T., Quay, H. C., & Conners, C. K. (1991). National survey of problems and competencies among four- to sixteen-year-olds: Parents' reports for normative and clinical samples. Monographs of the Society for Research in Child Development, 56 (3), v-130.

Achenbach, T. M., & McConaughy, S. H. (1987). Empirically based assessment of child and adolescent psychopathology: Practical applications (pp. 14-37, 94-99, 134-135, 140-162). Newbury Park, CA: Sage Publications.

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems implications of cross-informant correlations for situational specificity. Psychological Bulletin, 101 (2), 213-232.

Algozzine, B., Wong. B., & Obiakor, F. E. (1994). Children and youth with special needs, Diagnosis and classification of. In T. Husen & T. N. Postlethwaite (Eds.), International Encyclopedia of Education (pp. 711-719). City, NY: Pergamon Press.

Barnett, D. W. & Macmann, G. M. (1990). Personality assessment: Critical issues for research and practice. In C. R. Reynolds & R. W. Kamphaus (Eds.), Handbook of psychological and educational assessment of children (pp. 30-52). New York, NY: Guilford Press.

Barnett, D. W. & Zucker, K. B. (1990). The personal and social assessment of children: An analysis of current status and professional practice issues. Boston: Allyn and Bacon.

Bornstein, R. F., Rossner, S. C., Hill, E. L., & Stepanian, M. L. (1994). Face validity and fakability of objective and projective measures of dependency. Journal of Personality Assessment, 63 (1), 363-386.

Boyle, G. J., Stankov, L., & Cattell, R. B. (1995). Measurement and statistical models in the study of personality and intelligence. In D. H. Saklofske & M. Zeidner (eds.), International handbook of personality and intelligence (pp. 417-446). New York: Plenum Press.

Chen, W. I., Faraone, S. V., Riedferman, J., & Tsuang, M. T. (1994). Diagnostic accuracy of the Child Behavior Checklist scales for attention deficit hyperactivity disorder: A receiver-operating characteristic analysis. Journal of Consulting and Clinical Psychology, 62(5), 1017-1025.

De Groot, A., Koot, H. M. & Verhulst, F. C. (1994). Cross-cultural generalizability of the Child Behavior Checklist cross-informant syndromes. Psychological Assessment, 6 (3), 225-230.

Drotar, D., Stein, R. E. K., & Perrin, E. C. (1995). Methodological issues in using the Child Behavior Checklist and its related instruments in clinical child psychology research. Journal of Clinical Child Psychology, 24 (2), 184-192.

Elliot, S. N., Busse, R. T., & Gresham, F. M. (1993). Behavior rating scales: Issues of use and development. School Psychology Review, 22 (2), 313-321.

Emerson, E. N., Crowley, S. L., & Merrell, K. W. (1994). convergent validity of the School Social Behavior Scales with the Child Behavior Checklist and Teacher's Report Form. Journal of Psychoeducational Assessment, 12 (4), 372-380.

Epkins, C. C., & Meyers, A. W. (1994). Assessment of childhood depression, anxiety, and aggression: Convergent and discriminant validity of self-, parent-, teacher-, and peer-report measures. Journal of Personality Assessment, 62 (2), 364-381.

Fonhomme, E. (1992). Parent reports on behaviour and competencies among 6-11 year old French children. European Child and Adolescent Psychiatry, 1(4), 233-243.

Frankel, F., & Myatt, R. (1994). A dimensional approach to the assessment of social competence in boys. Psychological Assessment, 6(3), 249-254.

Hill, R. B., Billingsley, A., Engram, E., Malson, M. R., Rubin, R. H., Stack, C. B., Stewart, J. B., & Teele, J. E. (1993). Research on the African-American family: A holistic perspective. Westport, CT: Auburn House.

Jensen, A. R. (1980). Validity and correlates of mental tests. In Bias in mental testing (Ch. 8, pp. 297-366). New York: Free Press.

Jensen, A. R. (1980). Bias in predictive validity: Empirical evidence. In Bias in mental testing (Ch. 10, pp. 465-516). New York: Free Press.

Jensen, P. S., Traylor, J., Xenakis, S. N., & Davis, H. (1988a). Child psychopathology rating scales and interrater agreement: I. Parents' gender and psychiatric symptoms. Journal of the American Academy of child and Adolescent Psychiatry, 27 (4), 442-450.

Jensen, P. S., Traylor, J., Xenakis, S. N., & Davis, H. (1988b). Child psychopathology rating scales and interrater agreement: II. Child and family characteristics. Journal of the American Academy of child and Adolescent Psychiatry, 27 (4),451-461.

Kelley, M. L. (1985). Review of the Child Behavior Checklist. In J. V. Mitchell (Ed.), The ninth mental measurements yearbook: Vol. I (pp. 301-303). Lincoln, NE: University of Nebraska Press.

Kramer, J. J., & Conoley, J. C. (Eds.). (1990). Child Behavior Checklist. In The supplement to the tenth mental measurements yearbook (pp. 35-40). Lincoln, NE: University of Nebraska Press.

Lee, S. W., Elliott, J., & Barbour, J. d. (1994). A comparison of cross-informant behavior ratings in school-based diagnosis. Behavioral Disorders, 19 (2), 87-97.

Macmann, G. M., Barnett, D. W., & Lopez, E. J. (1993). The Child Behavior Checklist/4-18 and related materials: Reliability and validity of syndromal assessment. School Psychology Review, 22 (2), 322-333.

McConaughy, S. H. (1993). Advances in empirically based assessment of children's behavioral and emotional problems. School Psychology Review, (22) 2, 285-307.

McConaughy, S. H., Achenbach, T. M., & Gent, C. L. (1988). Multiaxial empirically based assessment: Parent, teacher, observational, cognitive, and personality correlates of child behavior profile types for 6- to 11-year-old boys. Journal of Abnormal Child Psychology, 16 (5), 485-509.

McConaughy, S. H., Mattison, R. E., & Peterson, R. L. (1994). Behavioral/emotional problems of children with serious emotional disturbances and learning disabilities. School Psychology Review, 23 (1), 81-98.

Mooney, K. C. (1984). Child Behavior Checklist. In D. J. Keyser & R. C. Sweetland (Eds.), Test critiques (Vol. 1., pp. 168-184). Kansas City, MO: Test Corporation of America.

Moran, M. M. (1990). The problem of cultural bias in personality assessment. In C. R. Reynolds & R. W. Kamphaus (Eds.), Handbook of psychological and educational assessment of children (pp. 524-545). New York, NY: Guilford Press.

Moreland, K. L. (1990). Computer-assisted assessment of adolescent and child personality: What's available. In C. R. Reynolds & R. W. Kamphaus (Eds.), Handbook of psychological and educational assessment of children (pp. 519-523). New York, NY: Guilford Press.

Most, R. B. & Zeidner, M. (1995). Constructing personality and intelligence instruments. In D. H. Saklofske & M. Zeidner (eds.), International handbook of personality and intelligence (pp. 475-503). New York: Plenum Press.

Newman, I., & Newman, C. (1994). Conceptual statistics for beginners (2nd ed.). New York: University Press of America.

Perrin, E. C., Stein, R. E. K., & Drotar, D., (1991). Cautions in using the Child Behavior Checklist: Observations based on research about children with a chronic illness. Journal of Pediatric Psychology, 16 (4), 411-421.

Raadal, M., Milgrom, P., Cauce, A. M., & Mancl, L. (1994 September). Behavior problems in 5- to 11-year-old children from low-income families. Journal of the American Academy of Child and Adolescent Psychiatry, 33 (7), 1017-1025.

Reynolds, C. R. (1995). Test bias and the assessment of intelligence and personality. In D. H. Saklofske & M. Zeidner (eds.), International handbook of personality and intelligence (pp. 545-573). New York: Plenum Press.

Tittle, C. K. (1994). Test bias. In T. Husen & T. N. Postlethwaite (Eds.), The International Encyclopedia of Education: Vol. 11 (2nd ed., pp. 6315-6321). Oxford, England: Elsevier Science Ltd.

Verhulst, F. C., Koot, H. M., & Van der Ende, J. (1994). Differential predictive value of parents' and teachers' reports of children's problem behaviors: A longitudinal study. Journal of Abnormal Child Psychology, 22 (5), 531-546.

Wiersma, W. & Jurs, S. G. (1990). Educational measurement and testing (2nd ed.). Boston: Allyn and Bacon.

Witt, J. C., Heffer, R. W., & Pfeiffer, J. (1990). Structured rating scales: A review of self-report and informant rating processes, procedures, and issues. In C. R. Reynolds & R. W. Kamphaus (Eds.), Handbook of psychological and educational assessment of children (pp. 363-394). New York, NY: Guilford Press.